

Parte I - Atividades teóricas

Exercício 1

- a) $P(A^C) = 1 - P(A) = 1 - \frac{1}{3} = \frac{2}{3}$
- b) $P(A^C \cup B) = P(A^C) + P(B) - P(A^C B) = P(A^C) + P(B) - (P(B) - P(AB)) = P(A^C) + P(AB) = \frac{2}{3} + \frac{1}{6} = \frac{5}{6}$
- c) $P(A \cup B^C) = P(A) + P(B^C) - P(AB^C) = P(B^C) + P(AB) = (1 - P(B)) + P(AB) = (1 - \frac{1}{4}) + \frac{1}{6} = \frac{11}{12}$
- d) $P(AB^C) = P(A) + P(B^C) - P(A \cup B^C) = \frac{1}{3} + \frac{3}{4} - \frac{11}{12} = \frac{2}{12} = \frac{1}{6}$
- e) $P(A^C \cup B^C) = 1 - P(AB) = 1 - \frac{1}{6} = \frac{5}{6}$

Exercício 2

a)

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(\xi) d\xi = \int_{-\infty}^x \frac{1}{2} d\xi = \left[\frac{1}{2} \xi \right]_0^x = \frac{1}{2} x, \forall x \in [0, 2]$$

b) $E\{X\} = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^2 \frac{1}{2} x dx = \left[\frac{x^2}{4} \right]_0^2 = 1$

$$E\{X^2\} = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^2 \frac{1}{2} x^2 dx = \left[\frac{x^3}{6} \right]_0^2 = \frac{4}{3}$$

$$E\{X^3\} = \int_{-\infty}^{\infty} x^3 f_X(x) dx = \int_0^2 \frac{1}{2} x^3 dx = \left[\frac{x^4}{8} \right]_0^2 = 2$$

Exercício 3

- a) X_2 , pois quanto mais próximo da distribuição uniforme, mais difícil é acertar o resultado de um evento aleatório "chutando um valor", ou seja, a variável aleatória X_2 carrega mais informação que a variável aleatória X_1 , pois neste último eu poderia supor que o resultado será sempre 3 com uma taxa de acertos de 40% contra uma taxa de acertos de 25% para qualquer valor estimado para X_2 se não tivermos nenhuma informação *a priori*.

$$\text{b) } H(X_1) = - \sum_x p(X_1) \log_2[p(X_1)] = -[0,1(-3,32) + 0,2(-2,32) + 0,3(-1,74) + 0,4(-1,32)] = 1,85$$

$$H(X_2) = - \sum_x p(X_2) \log_2[p(X_2)] = -[0,25(-2) + 0,25(-2) + 0,25(-2) + 0,25(-2)] = 2$$

$$\text{c) } D(P_1||P_2) = \sum_x p(X_1) \log_2 \left[\frac{p(X_1)}{p(X_2)} \right] = 0,1(-1,32) + 0,2(-0,32) + 0,3(-0,26) + 0,4(-0,68) = -0,54$$

$$D(P_2||P_1) = \sum_x p(X_2) \log_2 \left[\frac{p(X_2)}{p(X_1)} \right] = 0,25(1,32 + 0,32 - 0,26 - 0,68) = 0,18$$

Exercício 4

$$\text{a) } \mu_{ML} = \underset{\mu}{\operatorname{argmax}} p(x|\mu) = \underset{\mu}{\operatorname{argmax}} \log[p(x|\mu)] = \underset{\mu}{\operatorname{argmax}} \frac{p(x\mu)}{p(\mu)} = x$$

$$\begin{aligned} \text{b) } \mu_{ML} &= \underset{\mu}{\operatorname{argmax}} p(\mathbf{x}|\mu) = \underset{\mu}{\operatorname{argmax}} \log[p(\mathbf{x}|\mu)] = \underset{\mu}{\operatorname{argmax}} \sum_{k=1}^N \log[p(x_k|\mu)] = \\ &\underset{\mu}{\operatorname{argmax}} \sum_{k=1}^N \log \left[\frac{p(x_k\mu)}{p(\mu)} \right] \end{aligned}$$

$$\text{c) } \mu_{ML} = \frac{1}{N} \sum_{k=1}^N x_k$$

Parte II - Atividade computacional

a) Solução fechada com MMQ

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} \quad \Phi = \Phi(\mathbf{X}) = \begin{bmatrix} 1 & \mathbf{x}_1 \\ 1 & \mathbf{x}_2 \\ & \vdots \\ 1 & \mathbf{x}_N \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

Modelo:

$$\hat{y} = \hat{y}(\mathbf{X}) = \Phi^T \mathbf{w} \quad (1)$$

Solução ótima com o método de mínimos múltiplos quadrados (MMQ):

$$\mathbf{w} = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{y} \quad (2)$$

Raiz quadrada do erro quadrático médio (RMSE) para o modelo treinado com o conjunto de treino:

$$RMSE_{train} = 15.3702 \quad RMSE_{test} = 14.2495$$

Resultado no conjunto de teste:

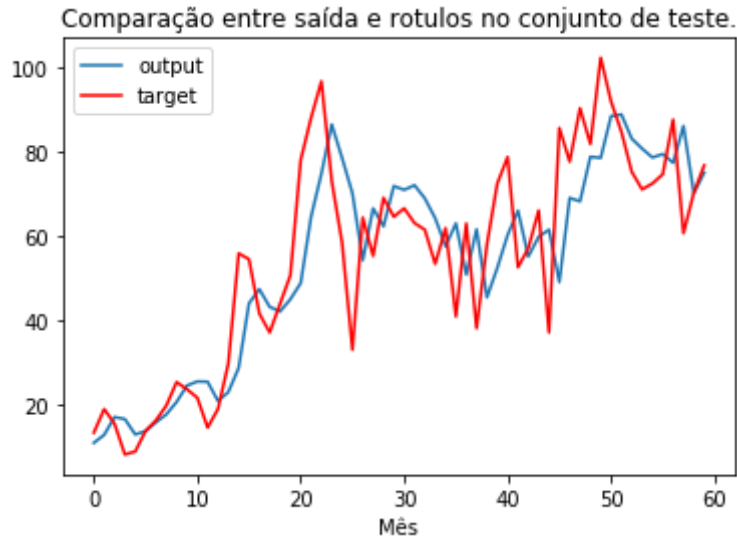


Figura 1: Saída prevista pelo modelo (azul) em comparação com valores reais dos últimos 5 anos (60 meses).

b) Seleção de atributos usando wrapper (backward elimination), validação cruzada (k-fold) e regularização L2

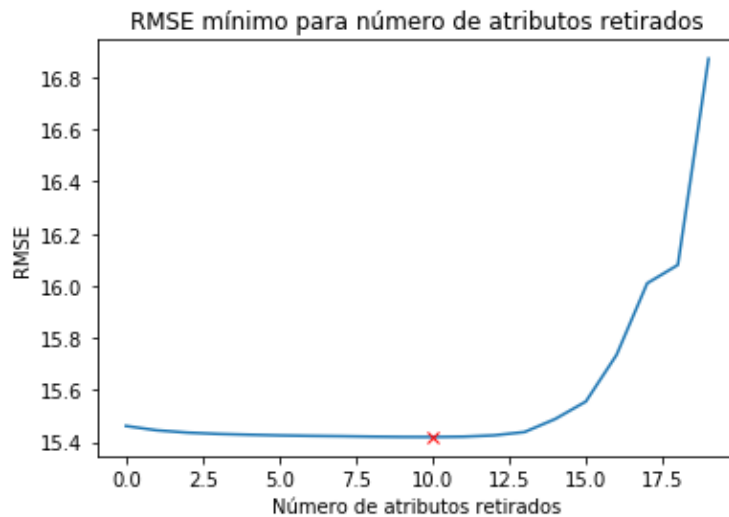


Figura 2: Menor RMSE para cada passo do wrapper (atributo retirado).

Modelo escolhido pelo wrapper utilizando k-fold com 5 pastas:

Número de atributos removidos: 10 λ : 70 RMSE: 14.1138

Atributos escolhidos: 0 (bias) , 1, 2, 3, 4, 5, 6, 9, 11, 18, 20

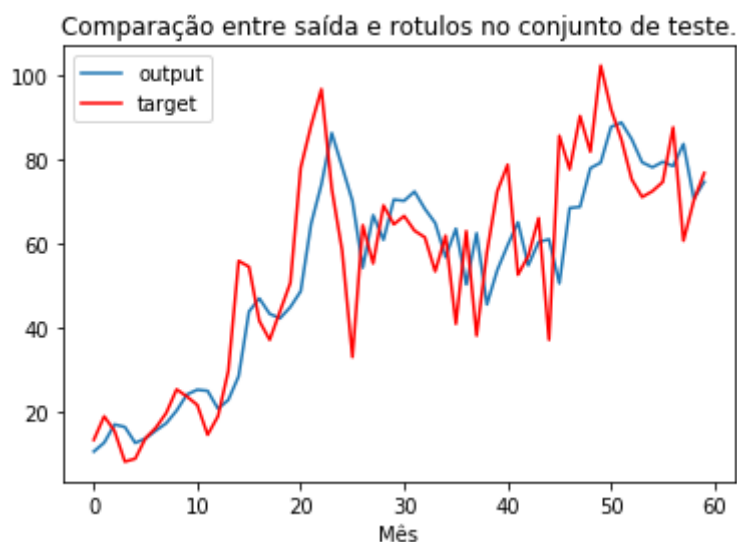


Figura 3: Saída prevista pelo modelo (azul) em comparação com valores reais dos últimos 5 anos (60 meses) com seleção de atributos usando wrapper.

c) Seleção de atributos usando filtro (correlação de Pearson)

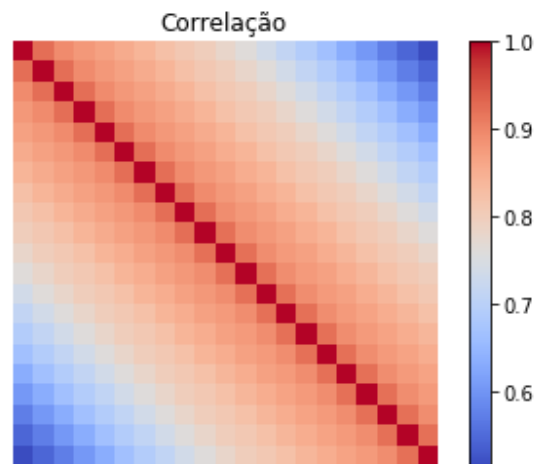


Figura 4: Matriz de correlação, onde uma cor mais avermelhada significa uma correlação maior (dados foram alinhados na matriz de forma que a primeira linha e coluna correspondem ao rótulo e as próximas linhas e colunas correspondem aos atributos - mês anterior até 20 meses atrás).

Atributos selecionados: 0 (bias) , 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

$\lambda : 30$ $RMSE : 13.8698$

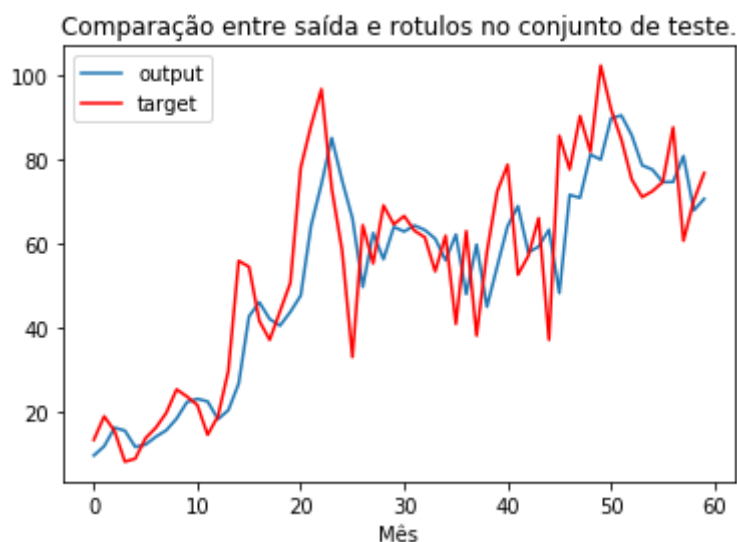


Figura 5: Saída prevista pelo modelo (azul) em comparação com valores reais dos últimos 5 anos (60 meses) com seleção de atributos usando filtro.

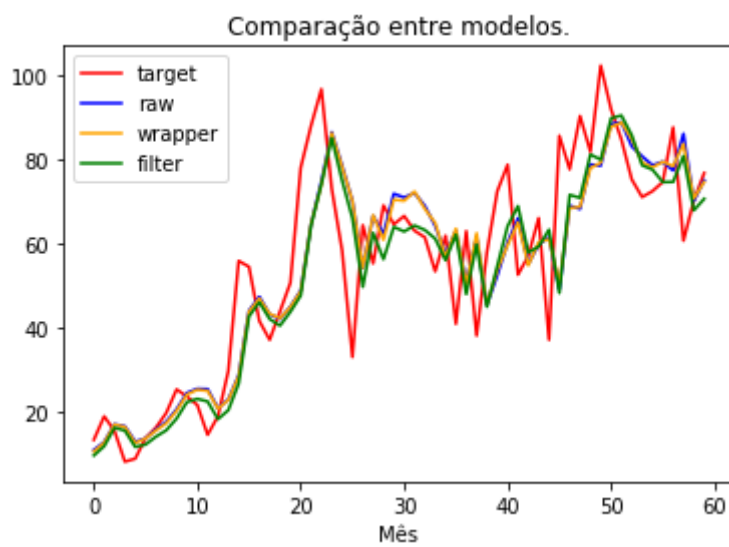


Figura 6: Saída prevista por cada modelo: azul - modelo inicial com 20 atributos, $RMSE = 14.2495$; amarelo - modelo criado usando wrapper, $RMSE = 14.1138$; verde - modelo criado usando filtro, $RMSE = 13.8698$; em comparação com valores reais (vermelho) dos últimos 5 anos (60 meses).